

# SIFT gegen SURF

## Essay

Technische Universität München  
May 18, 2015

### Abstract

Ohne funktionierende Algorithmen zu Merkmalsextraktion aus Bildern, würde in unserem heutigen Leben schon vieles nicht mehr so funktionieren wie wir es mittlerweile gewohnt sind. Dieses Essay beschäftigt sich mit der Frage welcher der State of the Art Algorithmen, SIFT oder SURF, besser geeignet ist im Hinblick auf Robustheit und Geschwindigkeit.

Es zeigt sich, dass SIFT zwar robuster ist, allerdings einen deutlichen Geschwindigkeitsnachteil gegenüber SURF aufweist. Gründe dafür sind die unterschiedlichen Verfahren zur Merkmalsdetektion. Im ersten Fall wird die Gaußdifferenz Funktion verwendet. SURF setzt hingegen auf Integralbilder mit konstanter Rechenzeit.

## Einführung

Merkmalsextraktion aus Bildern ist ein immer wichtiger werdendes Thema. In den meisten Smartphones finden sich bereits entsprechende Algorithmen. Google und viele andere digitale Unternehmen haben eine Vielzahl an Projekten, die sich mit dem Thema beschäftigen. Ein konkretes Beispiel ist die Bilderverwaltungssoftware Picassa, die Bilder nach bestimmten Ähnlichkeitsmerkmalen sortieren kann.

Die Algorithmen die hinter diesen Möglichkeiten stecken sind SIFT (Scale-invariant feature transform: *skaleninvariante Merkmalstransformation*) von David G. Lowe [1] und SURF (Speeded Up Robust Features: *Beschleunigte, robuste Merkmale*) von Bay, Ess, Tuytelaars und Van Gool [2].

Dieses Essay beschäftigt sich mit den Unterschieden und den damit verbundenen Vorteilen der beiden Algorithmen. Zuerst werden die technischen

Gegensätze beleuchtet und den damit verbundenen Varianzen in Qualität und Schnelligkeit. Anschließend wird jeweils eine geeignete Anwendungsmöglichkeit vorgestellt.

Meiner Meinung nach ist SURF zu bevorzugen. Hierbei zeigt sich eine höhere Geschwindigkeit als SIFT und ist so für viele, gerade mobile Anwendungen, besser geeignet. Diese Kernaussage wird im folgenden Kapitel erläutert.

## SIFT oder SURF

### SIFT

Der SIFT Algorithmus lässt sich in vier Schritte unterteilen:

1. Extremata Detektion
2. Lokalisation elementarer Punkte
3. Bestimmung der Orientierung
4. Beschreibung elementarer Punkte

Der entschiedenste Unterschied zu SURF offenbart sich bereits im ersten Teil, der Extremata Detektion. Es wird nach Kandidaten für möglichst sichere Merkmalspunkte gesucht, die unabhängig von Skalierungen des Bildes sind. Dafür wird das Bild kaskadenartig mit der Gaußdifferenz Funktion (*difference-of-Gaussian* - DoG) gefaltet.

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$

Jede Oktave wird auf diesem Weg verarbeitet und anschließend mit der halben Auflösung abgetastet. Somit reduziert sich der Rechenaufwand mit jeder Wiederholung.

Nach der Kandidatensuche, geht es um die exakte Lokalisierung dieser. Zum einen werden Kandidaten mit geringem Kontrast verworfen. Das geschieht mit Hilfe einer Taylor Approximation:

$$D(\hat{x}) = D + \frac{1}{2} \frac{dD^T}{dx} \hat{x}$$

Alle Punkte mit  $|D(\hat{x})|$  weniger als 0.03 werden verworfen.

Außerdem können über Berechnung der Determinante und Spur der Harris Matrix  $H$  von  $D$ , weitere instabile Extremata ausgeschlossen werden:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r}$$

Als nächstes wird die Orientierung jedes elementaren Punkts bestimmt. Dafür wird die dominierende Richtung der Gradienten aus den approximierten Spitzen eines Histogramms über diese extrahiert.

Um diese Punkte nun anschließend noch genauer zu beschreiben werden Gruppen von Histogrammen der umgebenden Pixel generiert. Dieser wiederum bestehen aus Untergruppen von 4x4 Pixeln. Gewichtet mit einer Gaußfunktion und schließlich normalisiert, entsteht so ein beschreibender Vektor.

## SURF

Der SURF Algorithmus ist dem SIFT sehr ähnlich. Auch dieser Algorithmus lässt sich in mehrere Phasen einteilen. Als erstes geht es um das Auffinden von interessanten Punkten. Das geschieht bei SURF allerdings mit Hilfe von Integralbildern:

$$I_{\Sigma}(x) \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j)$$

Der große Vorteil gegenüber dem Gaußfilter bei SIFT ist die konstante Zeit mit der das Integralbild berechnet werden kann, da diese unabhängig von der Filtergröße ist.

Ab diesem Punkt verfahren beide Algorithmen sehr ähnlich. SURF verarbeitet ebenfalls den Skalierungsbereich in mehrere Oktaven. Die Filtergröße wird dabei jeweils angepasst, was sich allerdings nicht auf die Rechenzeit auswirkt.

Für die Lokalisierung wird nun das gleiche Verfahren wie bei SIFT angewendet und erzeugt keine Veränderung in der Rechenzeit.

Ein weiterer Unterschied zu SIFT tritt bei der Beschreibung der Merkmalspunkte auf. SURF nutzt hier ein vereinfachtes Verfahren, welches mithilfe von Haar Matrizen als Filter, die Gradienteninformation extrahiert.

## Fazit

In [3] und [4] wurden Implementierungen beide Verfahren ausführlich getestet. Die Ergebnisse lassen sich wie folgt zusammenfassen:

SIFT hat einen Vorteil was die Robustheit anbelangt. Gerade unter schwierigen Bedingungen, wie zum Beispiel wenig Licht und geringer Kontrast, zeigt SIFT seine Stärken. Außerdem bietet es durchschnittlich mehr Merkmale pro Bild an als sein Gegenpart.

SURF hat hingegen einen signifikanten Geschwindigkeitsvorteil. Die Optimierungen durch die Integralbildmethode und der Haar Wavelets, führen zu einer besseren Rechenzeit. Daraus erwächst allerdings der Nachteil, dass

die Robustheit abnimmt. SURF kann ein Problem mit schwieriger Beleuchtung haben. Außerdem produziert es eine geringere Anzahl an Merkmalen.

Wenn man nun die sich daraus sinnvoll ergebenden Anwendungsgebiete der beiden Algorithmen betrachtet, kann eine geringere Zahl an Merkmalen auch ein Vorteil sein.

Besonders für den mobilen Bereich erscheint SURF sehr geeignet. Zum einen haben mobile Geräte noch nicht so viel Rechenkraft wie Desktop PCs, außerdem liefern viele Smartphonekameras noch keine qualitativ hochwertigen Bilder. Geschwindigkeit ist hingegen sehr entscheidend. Zum Beispiel möchte ein Nutzer der Augmented-Reality-Brille HoloLens von Microsoft kein Unterbrechungen feststellen, während ihm die Brille Informationen zu seiner Umgebung anzeigt.

SIFT kann seine Stärken dagegen ausspielen, wenn es um die offline Analyse von Bildmaterial geht. Hier stellt der Zeitfaktor kein all zu großes Problem mehr da. Das Augenmerk liegt hingegen auf der Qualität. Ein Beispiel ist das Stitching von Panoramabildern. Hierbei müssen sich überlappende Bilder, aus unterschiedlichen Richtungen aufgenommen, zu einem großen Bild zusammen gesetzt werden. Eine hohe Genauigkeit ist dabei wichtig, um störende Übergänge zu vermeiden.

## References

1. David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
2. Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.
3. Johannes Bauer, Niko Sunderhauf, and Peter Protzel. Comparing several implementations of two recently published feature detectors. In *Proc. of the International Conference on Intelligent and Autonomous Systems*, volume 6, 2007.
4. Edouard Oyallon and Julien Rabin. An analysis and implementation of the surf method, and its comparison to sift. *Image Processing On Line*, pages 1–31, 2013.